EA2020-descriptive.R

Edre MA, DrPH

2020-12-02

```
# ======================
# Descriptive Statistics
# R Biostat Workshop IIUM
# Edre MA, DrPH
# ======================

#libraries needed to be installed

#foreign
#epiDisplay
#psych
#ggubr
#usingR

# data

#pulling the data from GitHub

#go to https://github.com/adilzainal/IIUM_Biostatistic_workshop
#click "code" -> "Download ZIP"
#extract the ZIP file using WinRAR
#Create a new specific folder to store all files in your desktop
#set as working directory

#loading the data

#if spss (.sav)
library(foreign)
healthstat = read.spss("healthstatus.sav", to.data.frame = TRUE)

## re-encoding from UTF-8

str(healthstat)

## 'data.frame':    153 obs. of  12 variables:
##  $ id      : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ age     : num  36 49 56 61 40 42 44 41 46 32 ...
##  $ sex     : Factor w/ 2 levels "Female","Male": 2 2 2 1 1 1 2 2 1 1 ...
##  $ exercise: Factor w/ 3 levels "Low","Moderate",..: 2 1 1 2 2 2 3 1 3 1
...
##  $ smoking : Factor w/ 2 levels "No","Yes": 2 2 2 1 1 1 2 2 2 1 ...
##  $ wt      : num  58.5 64.7 63 47.4 44.8 58.9 56.4 46.1 70 81.1 ...
##  $ ht      : num  145 166 145 158 150 144 162 147 163 167 ...
```

```
##  $ sbp     : num  120 123 125 131 116 149 117 148 121 119 ...
##  $ dbp     : num  86 106 103 87 88 99 83 112 85 88 ...
##  $ hba1c   : num  10.1 7.2 8.7 6.9 5.6 10.3 5.9 10.1 3.5 5.4 ...
##  $ hcy     : num  4.78 11.18 8.65 6.2 5.36 ...
##  $ wt2     : num  49.5 62.2 62.2 43 40 53.9 54.3 43 68.2 79.4 ...
##  - attr(*, "variable.labels")= Named chr  "" "Age (years)" "Sex" "Exercise
intensity" ...
##   ..- attr(*, "names")= chr  "id" "age" "sex" "exercise" ...
##  - attr(*, "codepage")= int 65001
```

```
summary(healthstat)
```

```
##        id           age            sex         exercise  smoking
##  Min.   :  1   Min.   :21.00   Female:70   Low     :61   No :105
##  1st Qu.: 39   1st Qu.:36.00   Male  :83   Moderate:62   Yes: 48
##  Median : 77   Median :42.00               High    :30
##  Mean   : 77   Mean   :42.16
##  3rd Qu.:115   3rd Qu.:47.00
##  Max.   :153   Max.   :64.00
##        wt               ht             sbp             dbp
##  Min.   : 37.70   Min.   :140.0   Min.   : 99.0   Min.   : 69.00
##  1st Qu.: 50.60   1st Qu.:148.0   1st Qu.:116.0   1st Qu.: 83.00
##  Median : 58.90   Median :157.0   Median :122.0   Median : 88.00
##  Mean   : 61.68   Mean   :156.1   Mean   :123.9   Mean   : 90.37
##  3rd Qu.: 68.40   3rd Qu.:162.0   3rd Qu.:132.0   3rd Qu.: 97.00
##  Max.   :109.10   Max.   :176.0   Max.   :149.0   Max.   :123.00
##      hba1c             hcy              wt2
##  Min.   : 3.300   Min.   : 4.054   Min.   : 33.30
##  1st Qu.: 5.500   1st Qu.: 5.992   1st Qu.: 47.00
##  Median : 6.800   Median : 8.492   Median : 55.10
##  Mean   : 7.001   Mean   : 8.901   Mean   : 57.75
##  3rd Qu.: 8.500   3rd Qu.:10.622   3rd Qu.: 64.90
##  Max.   :11.600   Max.   :23.600   Max.   :107.60
```

```
View(healthstat)
```

```
#if excel (.xlsx)
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.3
```

```
healthstat <- read_excel("healthstatus.xlsx")
View(healthstat)
```

```
#summarising numerical values
```

```
# central tendency & dispersion
mean(healthstat$sbp)
```

```
## [1] 123.902
```

```
mean(healthstat$age)
```

```
## [1] 42.1634
```

```
sd(healthstat$sbp)
```

```
## [1] 11.31648
```

```
sd(healthstat$age)
```

```
## [1] 8.932096
```

```
median(healthstat$sbp)
```

```
## [1] 122
```

```
median(healthstat$age)
```

```
## [1] 42
```

```
IQR(healthstat$sbp)
```

```
## [1] 16
```

```
IQR(healthstat$age)
```

```
## [1] 11
```

```
# describe using sapply
mean_all = sapply(healthstat[c("age", "sbp", "dbp")], mean)
mean_all
```

```
##       age       sbp       dbp
##  42.16340 123.90196  90.36601
```

```
sd_all = sapply(healthstat[c("age", "sbp", "dbp")], sd)
sd_all
```

```
##       age       sbp       dbp
##  8.932096 11.316479 11.148962
```

```
median_all = sapply(healthstat[c("age", "sbp", "dbp")], median)
median_all
```

```
## age sbp dbp
##  42 122  88
```

```
iqr_all = sapply(healthstat[c("age", "sbp", "dbp")], IQR)
iqr_all
```

```
## age sbp dbp
##  11  16  14
```

```
cbind(Mean = mean_all, SD = sd_all, Median = median_all, IQR = iqr_all)
```

```
##          Mean        SD Median IQR
## age  42.16340  8.932096     42  11
```

```
## sbp 123.90196 11.316479    122  16
## dbp  90.36601 11.148962     88  14
```

```
rbind(Mean = mean_all, SD = sd_all, Median = median_all, IQR = iqr_all)
```

```
##               age       sbp      dbp
## Mean   42.163399 123.90196 90.36601
## SD      8.932096  11.31648 11.14896
## Median 42.000000 122.00000 88.00000
## IQR    11.000000  16.00000 14.00000
```

```
#normality assumption

#mean~median
#acceptable skewness & kurtosis +-2d
#bell shaped curve
#normality test

# describe using codebook, gives you mean~median
library(epiDisplay)
```

```
## Warning: package 'epiDisplay' was built under R version 3.6.3
```

```
## Loading required package: survival
```

```
## Loading required package: MASS
```

```
## Loading required package: nnet
```

```
codebook(healthstat)
```

```
##
##
##
## id    :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 id        153  77     77      44.31  1      153
##
##  ==================
## age   :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 age       153  42.16  42      8.93   21     64
##
##  ==================
## sex   :
```

```
## Warning in na.omit(as.numeric(x[[i]])): NAs introduced by coercion

##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 sex
##
##  ==================
## exercise      :

## Warning in na.omit(as.numeric(x[[i]])): NAs introduced by coercion

##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 exercise
##
##  ==================
## smoking   :

## Warning in na.omit(as.numeric(x[[i]])): NAs introduced by coercion

##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 smoking
##
##  ==================
## wt    :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 wt        153  61.68  58.9    15.06  37.7   109.1
##
##  ==================
## ht    :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 ht        153  156.09 157     8.81   140    176
##
##  ==================
## sbp   :
##
## No. of observations = 153
##
```

```
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 sbp       153  123.9  122     11.32  99     149
##
##  ==================
## dbp   :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 dbp       153  90.37  88      11.15  69     123
##
##  ==================
## hba1c     :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 hba1c     153  7      6.8     1.93   3.3    11.6
##
##  ==================
## hcy   :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 hcy       153  8.9    8.49    3.71   4.05   23.6
##
##  ==================
## wt2   :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 wt2       153  57.75  55.1    15.27  33.3   107.6
##
##  ==================
```

```
codebook(healthstat[c("age", "sbp", "dbp")])
```

```
##
##
##
## age   :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 age       153  42.16  42      8.93   21     64
##
##  ==================
```

```
## sbp   :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 sbp       153  123.9  122     11.32  99     149
##
##  ==================
## dbp   :
##
## No. of observations = 153
##
##   Var. name obs. mean   median  s.d.   min.   max.
## 1 dbp       153  90.37  88      11.15  69     123
##
##  ==================
```

```
# describe using describe, gives you skewness & kurtosis
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:epiDisplay':
##
##     alpha, cs, lookup
```

```
describe(healthstat[c("age", "sbp", "dbp")])
```

```
##     vars   n   mean    sd median trimmed   mad min max range skew kurtosis
se
## age    1 153  42.16  8.93     42   41.98  8.90  21  64    43 0.16    -0.25
0.72
## sbp    2 153 123.90 11.32    122  123.72 11.86  99 149    50 0.17    -0.60
0.91
## dbp    3 153  90.37 11.15     88   89.58  8.90  69 123    54 0.69    -0.02
0.90
```

```
# Determining normality of numerical data: bell shaped curve
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.6.3
```

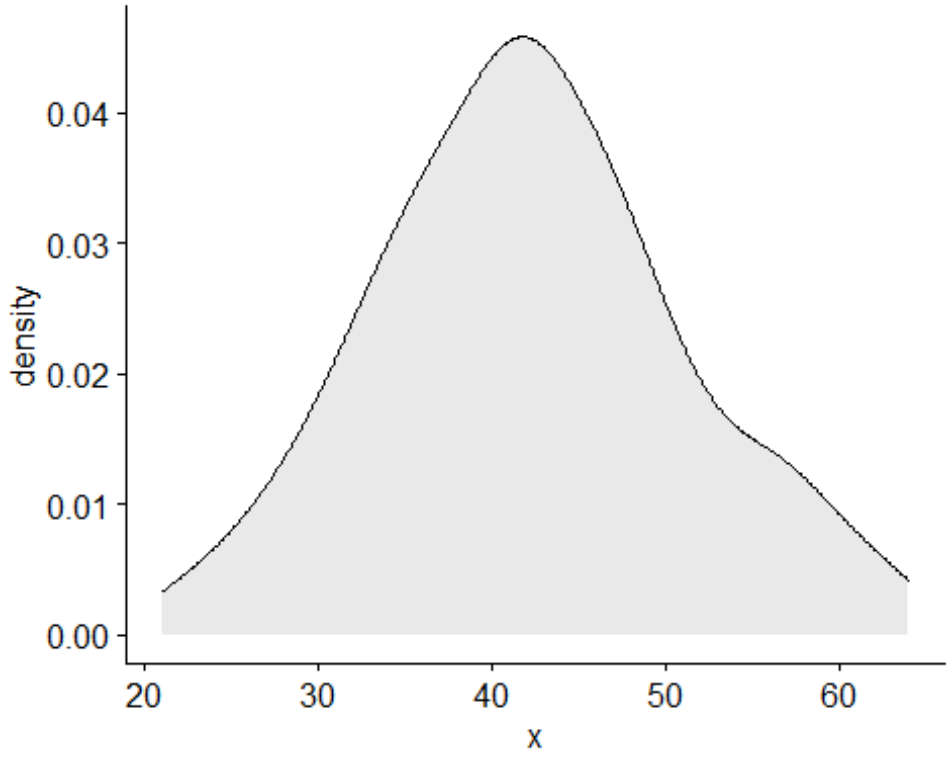
```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha

## The following object is masked from 'package:epiDisplay':
##
##     alpha
```

```
ggdensity(healthstat$age, fill = "lightgray")
```

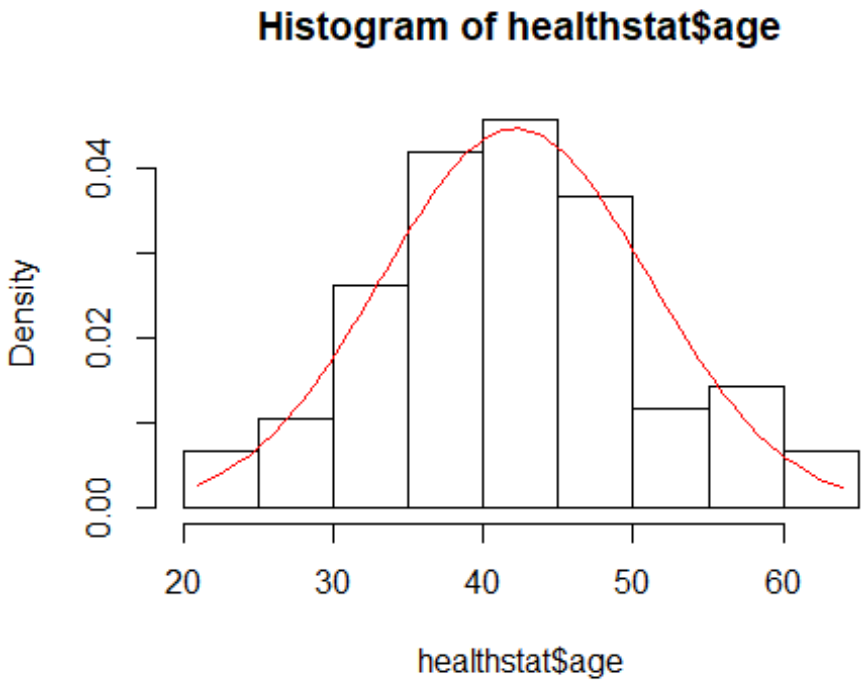



```
library(UsingR)
```

```
## Warning: package 'UsingR' was built under R version 3.6.3

## Loading required package: HistData

## Warning: package 'HistData' was built under R version 3.6.3

## Loading required package: Hmisc

## Warning: package 'Hmisc' was built under R version 3.6.3

## Loading required package: lattice

##
## Attaching package: 'lattice'
```

```
## The following object is masked from 'package:epiDisplay':
##
##     dotplot

## Loading required package: Formula

## Warning: package 'Formula' was built under R version 3.6.3

##
## Attaching package: 'Hmisc'

## The following object is masked from 'package:psych':
##
##     describe

## The following objects are masked from 'package:base':
##
##     format.pval, units

##
## Attaching package: 'UsingR'

## The following object is masked from 'package:psych':
##
##     headtail

## The following object is masked from 'package:survival':
##
##     cancer
```

```r
hist(healthstat$age, freq = FALSE)
x <- seq(21, 64, length.out=100)
y <- with(healthstat, dnorm(x, mean(age), sd(age)))
lines(x, y, col = "red")
```




```
# Determining normality of numerical data: normality test
shapiro.test(healthstat$age) #if data sample size is <50

##
##  Shapiro-Wilk normality test
##
## data:  healthstat$age
## W = 0.99149, p-value = 0.4934

#summarising categorical values

# proportion
tab_sex = table(healthstat$sex)
tab_smoking = table(healthstat$smoking)
tab_sex

##
## Female   Male
##     70     83

tab_smoking

##
##  No Yes
## 105  48

str(tab_sex)
```

```
##  'table' int [1:2(1d)] 70 83
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "Female" "Male"
```

```
str(tab_smoking)
```

```
##  'table' int [1:2(1d)] 105 48
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "No" "Yes"
```

```
prop.table(tab_sex)
```

```
##
##    Female      Male
## 0.4575163 0.5424837
```

```
prop.table(tab_smoking)
```

```
##
##        No       Yes
## 0.6862745 0.3137255
```

```
prop.table(tab_sex)*100
```

```
##
##   Female     Male
## 45.75163 54.24837
```

```
prop.table(tab_smoking)*100
```

```
##
##       No      Yes
## 68.62745 31.37255
```

```
#crosstabulation
smokingbygender<-table(healthstat$sex, healthstat$smoking)
prop.table(smokingbygender, margin=1)
```

```
##
##                 No       Yes
##   Female 0.8857143 0.1142857
##   Male   0.5180723 0.4819277
```

```
prop.table(smokingbygender, margin=1)*100
```

```
##
##                No      Yes
##   Female 88.57143 11.42857
##   Male   51.80723 48.19277
```

```
# by groups (Stratified by a categorical variable)
by(healthstat$age, healthstat$sex, mean)
```

```
## healthstat$sex: Female
## [1] 42.77143
## ------------------------------------------------------------
## healthstat$sex: Male
## [1] 41.6506
```

```
by(healthstat$age, healthstat$sex, sd)
```

```
## healthstat$sex: Female
## [1] 9.404241
## ------------------------------------------------------------
## healthstat$sex: Male
## [1] 8.537484
```

```
by(healthstat$age, healthstat$smoking, mean)
```

```
## healthstat$smoking: No
## [1] 41.94286
## ------------------------------------------------------------
## healthstat$smoking: Yes
## [1] 42.64583
```

```
by(healthstat$age, healthstat$smoking, sd)
```

```
## healthstat$smoking: No
## [1] 9.357051
## ------------------------------------------------------------
## healthstat$smoking: Yes
## [1] 7.995982
```

```
by(healthstat$age, healthstat$sex, median)
```

```
## healthstat$sex: Female
## [1] 42
## ------------------------------------------------------------
## healthstat$sex: Male
## [1] 42
```

```
by(healthstat$age, healthstat$sex, IQR)
```

```
## healthstat$sex: Female
## [1] 11
## ------------------------------------------------------------
## healthstat$sex: Male
## [1] 10
```

```
by(healthstat$age, healthstat$smoking, median)
```

```
## healthstat$smoking: No
## [1] 41
## ------------------------------------------------------------
```

```
## healthstat$smoking: Yes
## [1] 42.5
```

```
by(healthstat$age, healthstat$smoking, IQR)
```

```
## healthstat$smoking: No
## [1] 11
## ------------------------------------------------------------
## healthstat$smoking: Yes
## [1] 11
```

```
#missing data

#usually coded as "NA" in the dataset

is.na(healthstat)
```

```
##           id   age   sex exercise smoking    wt    ht   sbp   dbp hba1c
hcy
##   [1,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [2,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [3,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [4,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [5,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [6,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [7,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [8,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##   [9,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [10,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [11,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [12,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [13,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [14,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [15,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [16,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
```

```
##  [17,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [18,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [19,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [20,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [21,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [22,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [23,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [24,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [25,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [26,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [27,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [28,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [29,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [30,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [31,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [32,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [33,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [34,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [35,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [36,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [37,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [38,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [39,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [40,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [41,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
```

```
##  [42,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [43,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [44,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [45,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [46,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [47,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [48,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [49,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [50,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [51,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [52,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [53,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [54,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [55,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [56,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [57,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [58,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [59,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [60,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [61,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [62,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [63,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [64,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [65,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [66,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
```

```
##  [67,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [68,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [69,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [70,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [71,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [72,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [73,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [74,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [75,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [76,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [77,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [78,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [79,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [80,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [81,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [82,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [83,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [84,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [85,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [86,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [87,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [88,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [89,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [90,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [91,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
```

```
##  [92,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [93,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [94,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [95,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [96,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [97,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [98,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [99,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [100,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [101,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [102,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [103,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [104,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [105,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [106,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [107,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [108,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [109,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [110,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [111,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [112,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [113,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [114,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [115,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [116,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
```

```
## [117,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [118,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [119,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [120,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [121,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [122,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [123,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [124,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [125,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [126,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [127,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [128,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [129,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [130,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [131,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [132,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [133,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [134,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [135,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [136,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [137,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [138,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [139,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [140,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [141,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
```

```
## [142,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [143,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [144,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [145,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [146,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [147,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [148,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [149,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [150,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [151,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [152,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [153,] FALSE FALSE FALSE    FALSE   FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##          wt2
##   [1,] FALSE
##   [2,] FALSE
##   [3,] FALSE
##   [4,] FALSE
##   [5,] FALSE
##   [6,] FALSE
##   [7,] FALSE
##   [8,] FALSE
##   [9,] FALSE
##  [10,] FALSE
##  [11,] FALSE
##  [12,] FALSE
##  [13,] FALSE
##  [14,] FALSE
##  [15,] FALSE
##  [16,] FALSE
##  [17,] FALSE
##  [18,] FALSE
##  [19,] FALSE
##  [20,] FALSE
##  [21,] FALSE
##  [22,] FALSE
##  [23,] FALSE
##  [24,] FALSE
##  [25,] FALSE
```

```
##  [26,] FALSE
##  [27,] FALSE
##  [28,] FALSE
##  [29,] FALSE
##  [30,] FALSE
##  [31,] FALSE
##  [32,] FALSE
##  [33,] FALSE
##  [34,] FALSE
##  [35,] FALSE
##  [36,] FALSE
##  [37,] FALSE
##  [38,] FALSE
##  [39,] FALSE
##  [40,] FALSE
##  [41,] FALSE
##  [42,] FALSE
##  [43,] FALSE
##  [44,] FALSE
##  [45,] FALSE
##  [46,] FALSE
##  [47,] FALSE
##  [48,] FALSE
##  [49,] FALSE
##  [50,] FALSE
##  [51,] FALSE
##  [52,] FALSE
##  [53,] FALSE
##  [54,] FALSE
##  [55,] FALSE
##  [56,] FALSE
##  [57,] FALSE
##  [58,] FALSE
##  [59,] FALSE
##  [60,] FALSE
##  [61,] FALSE
##  [62,] FALSE
##  [63,] FALSE
##  [64,] FALSE
##  [65,] FALSE
##  [66,] FALSE
##  [67,] FALSE
##  [68,] FALSE
##  [69,] FALSE
##  [70,] FALSE
##  [71,] FALSE
##  [72,] FALSE
##  [73,] FALSE
##  [74,] FALSE
##  [75,] FALSE
```

```
##  [76,] FALSE
##  [77,] FALSE
##  [78,] FALSE
##  [79,] FALSE
##  [80,] FALSE
##  [81,] FALSE
##  [82,] FALSE
##  [83,] FALSE
##  [84,] FALSE
##  [85,] FALSE
##  [86,] FALSE
##  [87,] FALSE
##  [88,] FALSE
##  [89,] FALSE
##  [90,] FALSE
##  [91,] FALSE
##  [92,] FALSE
##  [93,] FALSE
##  [94,] FALSE
##  [95,] FALSE
##  [96,] FALSE
##  [97,] FALSE
##  [98,] FALSE
##  [99,] FALSE
## [100,] FALSE
## [101,] FALSE
## [102,] FALSE
## [103,] FALSE
## [104,] FALSE
## [105,] FALSE
## [106,] FALSE
## [107,] FALSE
## [108,] FALSE
## [109,] FALSE
## [110,] FALSE
## [111,] FALSE
## [112,] FALSE
## [113,] FALSE
## [114,] FALSE
## [115,] FALSE
## [116,] FALSE
## [117,] FALSE
## [118,] FALSE
## [119,] FALSE
## [120,] FALSE
## [121,] FALSE
## [122,] FALSE
## [123,] FALSE
## [124,] FALSE
## [125,] FALSE
```

```
## [126,] FALSE
## [127,] FALSE
## [128,] FALSE
## [129,] FALSE
## [130,] FALSE
## [131,] FALSE
## [132,] FALSE
## [133,] FALSE
## [134,] FALSE
## [135,] FALSE
## [136,] FALSE
## [137,] FALSE
## [138,] FALSE
## [139,] FALSE
## [140,] FALSE
## [141,] FALSE
## [142,] FALSE
## [143,] FALSE
## [144,] FALSE
## [145,] FALSE
## [146,] FALSE
## [147,] FALSE
## [148,] FALSE
## [149,] FALSE
## [150,] FALSE
## [151,] FALSE
## [152,] FALSE
## [153,] FALSE
```

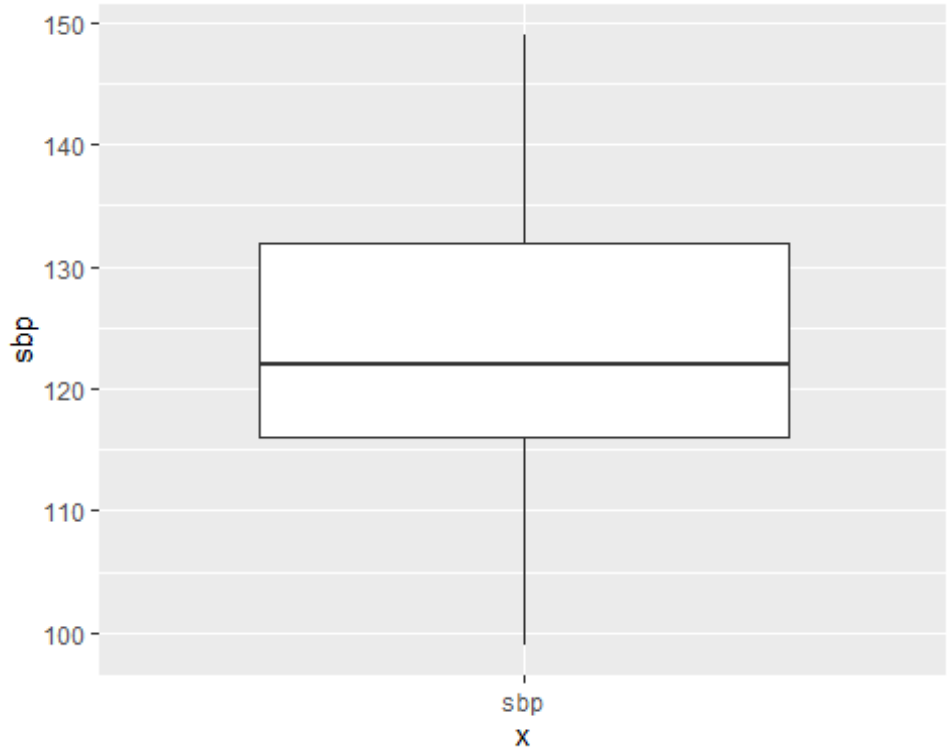
```
which (is.na(healthstat$sbp))
```

```
## integer(0)
```

```
#demonstrating the row to show the missing value using dummy data
x<- c(1,13,14,NA,2,44)
which (is.na(x))
```

```
## [1] 4
```

```
#outlier detection

#visual method
ggplot(healthstat, aes(x = "sbp", y = sbp)) + geom_boxplot()
```




```
#data row method

is_outlier <- healthstat$age > 150 | healthstat$age < 0
is_outlier
```

```
##   [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
##  [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
```

```
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

#basic data transformation:categorizing

#glucose control (6.5% and above considered poor)
healthstat$glucontrol<-cut(healthstat$hba1c, breaks=c(-
Inf,6.49,Inf),labels=c("good", "poor"))
summary(healthstat)

##        id           age            sex              exercise
##  Min.   :  1   Min.   :21.00   Length:153         Length:153
##  1st Qu.: 39   1st Qu.:36.00   Class :character   Class :character
##  Median : 77   Median :42.00   Mode  :character   Mode  :character
##  Mean   : 77   Mean   :42.16
##  3rd Qu.:115   3rd Qu.:47.00
##  Max.   :153   Max.   :64.00
##    smoking                wt               ht             sbp
##  Length:153         Min.   : 37.70   Min.   :140.0   Min.   : 99.0
##  Class :character   1st Qu.: 50.60   1st Qu.:148.0   1st Qu.:116.0
##  Mode  :character   Median : 58.90   Median :157.0   Median :122.0
##                     Mean   : 61.68   Mean   :156.1   Mean   :123.9
##                     3rd Qu.: 68.40   3rd Qu.:162.0   3rd Qu.:132.0
##                     Max.   :109.10   Max.   :176.0   Max.   :149.0
##       dbp             hba1c             hcy              wt2
glucontrol
##  Min.   : 69.00   Min.   : 3.300   Min.   : 4.054   Min.   : 33.30
good:60
##  1st Qu.: 83.00   1st Qu.: 5.500   1st Qu.: 5.992   1st Qu.: 47.00
poor:93
##  Median : 88.00   Median : 6.800   Median : 8.492   Median : 55.10
##  Mean   : 90.37   Mean   : 7.001   Mean   : 8.901   Mean   : 57.75
##  3rd Qu.: 97.00   3rd Qu.: 8.500   3rd Qu.:10.622   3rd Qu.: 64.90
##  Max.   :123.00   Max.   :11.600   Max.   :23.600   Max.   :107.60

#bmistatus (WHO classification)
healthstat$bmi <- (healthstat$wt)/(healthstat$ht/100)**2
healthstat$bmi

##   [1] 27.82402 23.47946 29.96433 18.98734 19.91111 28.40471 21.49063
21.33370
##   [9] 26.34649 29.07957 23.48596 29.04783 19.60440 17.86573 34.24772
18.43611
##  [17] 21.69625 31.75352 29.12194 44.05588 34.00402 28.96552 18.54031
22.18990
##  [25] 24.14062 24.17948 27.13500 25.35084 23.84236 24.20790 26.03678
19.01387
##  [33] 19.50379 21.51881 33.53147 27.42511 24.66632 22.29938 20.83000
22.15190
##  [41] 23.75155 27.53800 20.91552 35.22451 21.48437 30.91403 31.60551
```

```
25.49346
##  [49] 28.69964 28.55727 27.30104 26.02014 21.15247 32.48514 31.22717
28.57875
##  [57] 22.70168 28.75677 28.31123 18.35849 18.36727 19.73815 27.49014
27.14158
##  [65] 35.84775 33.62428 28.50116 22.80990 20.06243 25.18671 24.90973
18.51852
##  [73] 29.09469 29.25310 27.23922 20.97503 22.57778 19.39227 32.64244
27.96053
##  [81] 19.68750 29.70679 25.60554 22.75556 20.19558 33.15894 26.56434
20.73722
##  [89] 20.47499 20.42242 24.01013 24.10236 22.46845 20.64516 30.36885
19.61433
##  [97] 20.73001 23.82222 18.78463 24.62473 21.18335 19.65866 24.70588
24.68769
## [105] 35.85601 30.86801 31.98179 23.55734 21.82644 23.67409 36.35117
22.11863
## [113] 20.01503 23.83432 27.63894 26.76051 24.40000 20.32537 18.52237
38.10976
## [121] 30.37649 33.15644 19.67677 25.61176 30.70312 30.61224 18.39890
33.12783
## [129] 18.32800 45.41103 17.80270 23.66864 30.00000 24.84694 19.80584
27.11250
## [137] 27.49109 26.57778 23.43750 26.94384 21.06631 21.12573 22.60146
23.63237
## [145] 17.94584 20.51913 18.50796 20.21527 20.06920 29.19188 27.39726
18.95317
## [153] 25.23051
```

```
healthstat$bmistat <- cut(healthstat$bmi, breaks=c(-Inf, 18.49999, 24.9999,
29.9999, Inf), labels=c("underweight", "normal", "overweight", "obese"))
healthstat$bmistat
```

```
##   [1] overweight  normal      overweight  normal      normal
overweight
##   [7] normal      normal      overweight  overweight  normal
overweight
##  [13] normal      underweight obese       underweight normal      obese
##  [19] overweight  obese       obese       overweight  normal      normal
##  [25] normal      normal      overweight  overweight  normal      normal
##  [31] overweight  normal      normal      normal      obese
overweight
##  [37] normal      normal      normal      normal      normal
overweight
##  [43] normal      obese       normal      obese       obese
overweight
##  [49] overweight  overweight  overweight  overweight  normal      obese
##  [55] obese       overweight  normal      overweight  overweight
underweight
##  [61] underweight normal      overweight  overweight  obese       obese
```

```
##  [67] overweight  normal      normal      overweight  normal      normal
##  [73] overweight  overweight  overweight  normal      normal      normal
##  [79] obese       overweight  normal      overweight  overweight  normal
##  [85] normal      obese       overweight  normal      normal      normal
##  [91] normal      normal      normal      normal      obese       normal
##  [97] normal      normal      normal      normal      normal      normal
## [103] normal      normal      obese       obese       obese       normal
## [109] normal      normal      obese       normal      normal      normal
## [115] overweight  overweight  normal      normal      normal      obese
## [121] obese       obese       normal      overweight  obese       obese
## [127] underweight obese       underweight obese       underweight normal
## [133] obese       normal      normal      overweight  overweight
overweight
## [139] normal      overweight  normal      normal      normal      normal
## [145] underweight normal      normal      normal      normal
overweight
## [151] overweight  normal      overweight
## Levels: underweight normal overweight obese
```

```
#hypertension status (either sbp or dbp equal or more than 140/90mmHg,
respectively, considered hypertensive)
healthstat$hpt<-(healthstat$sbp>=140|healthstat$dbp>=90)
summary(healthstat) #logical class for the new outcome
```

```
##        id           age            sex              exercise
##  Min.   :  1   Min.   :21.00   Length:153         Length:153
##  1st Qu.: 39   1st Qu.:36.00   Class :character   Class :character
##  Median : 77   Median :42.00   Mode  :character   Mode  :character
##  Mean   : 77   Mean   :42.16
##  3rd Qu.:115   3rd Qu.:47.00
##  Max.   :153   Max.   :64.00
##    smoking                wt               ht             sbp
##  Length:153         Min.   : 37.70   Min.   :140.0   Min.   : 99.0
##  Class :character   1st Qu.: 50.60   1st Qu.:148.0   1st Qu.:116.0
##  Mode  :character   Median : 58.90   Median :157.0   Median :122.0
##                     Mean   : 61.68   Mean   :156.1   Mean   :123.9
##                     3rd Qu.: 68.40   3rd Qu.:162.0   3rd Qu.:132.0
##                     Max.   :109.10   Max.   :176.0   Max.   :149.0
##       dbp             hba1c             hcy              wt2
glucontrol
##  Min.   : 69.00   Min.   : 3.300   Min.   : 4.054   Min.   : 33.30
good:60
##  1st Qu.: 83.00   1st Qu.: 5.500   1st Qu.: 5.992   1st Qu.: 47.00
poor:93
##  Median : 88.00   Median : 6.800   Median : 8.492   Median : 55.10
##  Mean   : 90.37   Mean   : 7.001   Mean   : 8.901   Mean   : 57.75
##  3rd Qu.: 97.00   3rd Qu.: 8.500   3rd Qu.:10.622   3rd Qu.: 64.90
##  Max.   :123.00   Max.   :11.600   Max.   :23.600   Max.   :107.60
##       bmi               bmistat      hpt
##  Min.   :17.80   underweight: 8   Mode :logical
```

```
##  1st Qu.:20.83   normal     :76   FALSE:88
##  Median :24.21   overweight :42   TRUE :65
##  Mean   :25.23   obese      :27
##  3rd Qu.:28.58
##  Max.   :45.41

healthstat$hpt2 <- as.factor(healthstat$hpt) #convert from logical to a
factor variable
summary(healthstat)

##        id           age            sex              exercise
##  Min.   :  1   Min.   :21.00   Length:153         Length:153
##  1st Qu.: 39   1st Qu.:36.00   Class :character   Class :character
##  Median : 77   Median :42.00   Mode  :character   Mode  :character
##  Mean   : 77   Mean   :42.16
##  3rd Qu.:115   3rd Qu.:47.00
##  Max.   :153   Max.   :64.00
##    smoking                wt               ht             sbp
##  Length:153         Min.   : 37.70   Min.   :140.0   Min.   : 99.0
##  Class :character   1st Qu.: 50.60   1st Qu.:148.0   1st Qu.:116.0
##  Mode  :character   Median : 58.90   Median :157.0   Median :122.0
##                     Mean   : 61.68   Mean   :156.1   Mean   :123.9
##                     3rd Qu.: 68.40   3rd Qu.:162.0   3rd Qu.:132.0
##                     Max.   :109.10   Max.   :176.0   Max.   :149.0
##       dbp             hba1c             hcy              wt2
glucontrol
##  Min.   : 69.00   Min.   : 3.300   Min.   : 4.054   Min.   : 33.30
good:60
##  1st Qu.: 83.00   1st Qu.: 5.500   1st Qu.: 5.992   1st Qu.: 47.00
poor:93
##  Median : 88.00   Median : 6.800   Median : 8.492   Median : 55.10
##  Mean   : 90.37   Mean   : 7.001   Mean   : 8.901   Mean   : 57.75
##  3rd Qu.: 97.00   3rd Qu.: 8.500   3rd Qu.:10.622   3rd Qu.: 64.90
##  Max.   :123.00   Max.   :11.600   Max.   :23.600   Max.   :107.60
##       bmi               bmistat      hpt             hpt2
##  Min.   :17.80   underweight: 8   Mode :logical   FALSE:88
##  1st Qu.:20.83   normal     :76   FALSE:88        TRUE :65
##  Median :24.21   overweight :42   TRUE :65
##  Mean   :25.23   obese      :27
##  3rd Qu.:28.58
##  Max.   :45.41

#Acknowledgement : Dr WNAriffin (USM)
```